Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 45 (2005) 2943–2959

Vision Research

www.elsevier.com/locate/visres

Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination?

M. Chirimuuta [a,b], D.J. Tolhurst [a,*]

[a] *Department of Physiology, Downing Street, Cambridge CB2 3EG, UK*
[b] *Istituto di Neurofisiologia CNR, Via Moruzzi, 1, Pisa 56100, Italy*

Received 16 July 2004; received in revised form 6 June 2005

Abstract

The dipper effect for contrast discrimination provides strong evidence that the underlying neural response is accelerating at low contrasts and saturating at high contrasts. The contrast–response functions of V1 neurons do have this sigmoidal shape, but individual neurons do not generally have a dynamic range wide enough to account for the dipper effect. This paper presents a Bayesian model of neurons in monkey V1, whose contrast–response function is described by a modified Naka–Rushton with multiplicative noise. It is shown that a model of groups of twelve or more neurons gives a reasonable explanation of the psychophysical data of two observers, but there is a large systematic error which is apparently due to the shape of the distribution of the monkey's sensitivity parameter, c_{50}. A further model provides a better fit to the data by sacrificing strict adherence to V1 neuronal parameters and, instead using an arbitrary bimodal c_{50} distribution, perhaps reflecting differences between M- and P-cells.

Keywords: Contrast coding; Dipper function; Bayesian model; Gratings; Contrast discrimination; Psychophysics; Modelling; V1

1. Introduction

The so-called "dipper function" is the well-documented result for psychophysical contrast discrimination: the threshold for detecting a contrast increment on a low-contrast *pedestal* (or *mask*) drops below the absolute detection threshold (*facilitation* effect), while high-contrast pedestals cause threshold elevation (*masking* effect); the graph of threshold increment contrast versus mask contrast is dipper shaped on double logarithmic axes (Bradley & Ohzawa, 1986; Campbell & Kulikowski, 1966; Foley, 1994; Itti, Koch, & Braun, 2000; Legge & Foley, 1980; Nachmias & Sansbury, 1974; Tolhurst & Barfield, 1978). The facilitatory dip at very low masking contrasts is said to be due to a positively accelerating, "non-linear contrast transducer function" (Nachmias & Sansbury, 1974), roughly equivalent to "subthreshold summation" (Campbell & Kulikowski, 1966; Tolhurst & Barfield, 1978). The Weber–Fechner like masking at higher masking contrasts has generally been attributed to a compressive nonlinearity at high contrasts (Gorea & Sagi, 2001; Legge & Foley, 1980; Nachmias & Kocher, 1970). This is illustrated schematically in Fig. 1, which shows the contrast–response function of the transducer of Legge and Foley (1980). While the transducer is accelerating at low contrasts, little difference in input contrast (ΔC) is required to give a criterion difference in response (ΔR), i.e., the discrimination threshold is minimal. As the curve decelerates at high contrasts, ΔC increases for a fixed ΔR, leading to higher thresholds. Note that the low-contrast acceleration of the sigmoidal function is almost equivalent to a hard threshold in an otherwise-linear function.

fMRI studies have suggested that human contrast discrimination and the dipper function are compatible with the way in which the amplitude of the BOLD signal

* Corresponding author. Tel.: +44 1223 333 889; fax: +44 1223 333 840.
E-mail address: djt12@cam.ac.uk (D.J. Tolhurst).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.06.022

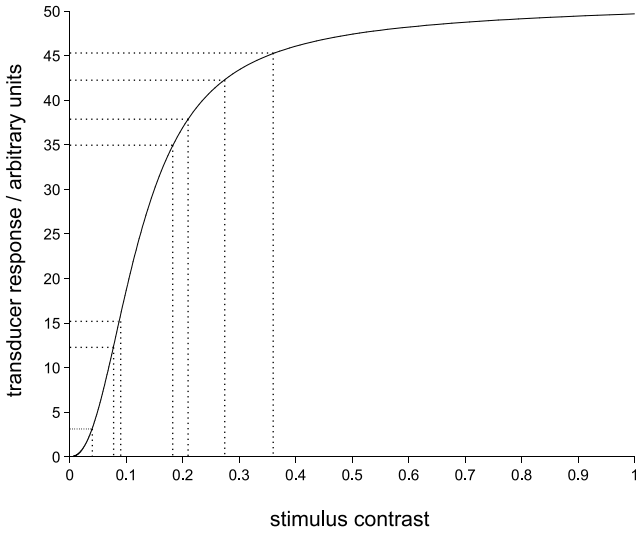

Fig. 1. The proposed non-linear transducer function (Legge & Foley, 1980). Response (arbitrary units) is plotted against contrast. Horizontal dotted lines mark off sections of fixed ΔR (hypothesised criterion for discrimination response) and the corresponding vertical dotted lines show how ΔC (contrast for threshold discrimination) varies with contrast level: ΔC first decreases as contrast increases from zero, then shows a large increase at high contrasts. This gives a dipping contrast discrimination curve. Note that the transducer has a wide dynamic range, since contrast is plotted on a linear axis.

changes with contrast in early visual areas such as primary visual cortex, V1 (Boynton, Demb, Glover, & Heeger, 1999; Heeger, Huk, Geisler, & Albrecht, 2000; Ress & Heeger, 2003). It is indeed true that the contrast–response functions of individual V1 neurons are sigmoidal, with positive acceleration at low contrasts and compression or saturation at high contrasts (Albrecht & Hamilton, 1982; Tolhurst, Movshon, & Thompson, 1981; Tolhurst, Movshon, & Dean, 1983). However, Itti et al. (2000) point out that the response variance or "noise" of V1 neurons increases with increasing contrast (Dean, 1981; Tolhurst et al., 1981, 1983; Vogels, Spileers, & Orban, 1989; Wiener, Oram, Liu, & Richmond, 2001), a phenomenon shown psychophysically by Kontsevich, Chen, and Tyler (2002). Increased noise at high contrast must increase the discrimination threshold, even if the transducer function were linear. Thus, a neurophysiological explanation of the dipper should not rely solely on considering how response *amplitude* depends upon contrast.

Furthermore, although individual V1 neurons *do* have sigmoidal response functions, these are not of the correct form to explain contrast discrimination over the wide range of visible contrasts, since any one neuron generally has limited dynamic range (Albrecht & Hamilton, 1982; Tolhurst et al., 1981, 1983). However, different neurons have their dynamic ranges in different parts of the visible contrast range. Heeger et al. (2000) noted that the fMRI BOLD signal was similar to the *summed* responses of many V1 neurons, covering different parts of the overall range. Thus, a neurophysiological model of human contrast discrimination must account for three features of neuronal responses: (i) each neuron has a sigmoidal response function, (ii) the dynamic range of most neurons is narrow and the overall contrast range is covered by a population of neurons with different dynamic ranges, (iii) response variance increases with increasing mean response.

In this paper, we develop a model of V1 encoding that encompasses these features (Chirimuuta, Clatworthy, & Tolhurst, 2003; Clatworthy, Chirimuuta, Lauritzen, & Tolhurst, 2003). It seems natural to employ a Bayesian approach, since this formalism allows one to optimally infer the value of a stimulus property when the link between stimulus and response is confounded by noise (Geisler & Albrecht, 1995, 1997), and it easily accommodates the need to model a population of neurons which have multiplicative noise. This is not to imply that the brain actually performs the Bayesian calculations; other decision rules are possible. The usefulness of the Bayesian framework is that it allows one to calculate what an ideal system might know, given the noisy information incoming from the sensory periphery (e.g., Rao, Olshausen, & Lewicki, 2002).

To simulate the contrast discrimination experiment, the model proceeds by first estimating the contrasts of the two stimulus intervals, then choosing the interval with the larger identified contrast. According to Bayes theorem (Eq. (1)), accurate contrast identification can be achieved if one infers that the contrast presented is the one for which the probability of contrast given response (the posterior probability, $P(c|r)$) is maximum, where,

$$P(c|r) = \frac{P(r|c) \cdot P(c)}{P(r)} \tag{1}$$

$P(r|c)$, the probability of response given contrast, is known as the *likelihood distribution*. $P(c)$ is known as the *prior* distribution and $P(r)$ is a normalising term.

Here, we model psychophysical contrast *discrimination* experiments (the "dipper"). The accompanying paper (Chirimuuta & Tolhurst, 2005) applies the same modelling to psychophysical contrast *identification*. Some of this work has been reported briefly (Chirimuuta, Clatworthy, & Tolhurst, 2002).

2. Methods

2.1. Psychophysical methods

2.1.1. Apparatus and stimuli

Grey-level stimuli were presented on a SONY 19 in. colour monitor driven by a VSG 2/4 graphics card (Cambridge Research Systems). Observers sat in a dimly lit room at a distance of 2.28 m from the screen, which

was 9.25 deg (37 cm) wide × 7 deg (28 cm) high. Viewing was binocular, with free fixation. The screen had a space-averaged mean luminance of 44 cd m^{-2}, bright enough to be in the photopic range.

The stimuli were vertical, 2.67 c deg^{-1} sinusoidal gratings and Gabor patches. These were all calculated as 256 × 256 pixels (where pixel size was 1.44 minutes of arc), represented to 256 grey levels, giving a maximal image size of 24 cm × 24 cm (6 deg × 6 deg at the viewing distance). The VSG 2/4 had "pseudo-15-bit" control of pixel luminance (Pelli & Zhang, 1991); this allowed correction of first-order luminance nonlinearities in the display while still allowing even the lowest contrast stimuli to be displayed with 256 grey levels. Each stimulus contrast was achieved by mapping the pixel value (8 bits) into a look-up table that had 256 values chosen from a palette of 2^{15}. The gratings had a Gaussian-weighted edge so that a there was no sharp border which might cue discrimination; this resulted in reduction in the size of the visible image by approximately the width of two cycles of the sinusoid (where the full-sized grating contains 16 cycles of the sinusoid). The Gabor patches had a spread of 16 pixels (0.38 deg), and the Gaussian envelope was calculated as

$$\text{weight}(x,y) = \exp\left(-\frac{(x^2+y^2)}{2\times \text{spread}^2}\right). \quad (2)$$

Michelson contrast of a grating is defined conventionally:

$$c = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}, \quad (3)$$

where $L_{\max}$ and $L_{\min}$ are the brightest and darkest pixels in a sinusoidal grating. It is convenient to refer to contrast as "dB attenuation from the maximum contrast of 1.0":

$$\text{dB} = -20\log_{10}(c). \quad (4)$$

The task in these experiments was to discriminate a *mask* stimulus from a composite *mask*-and-*test* stimulus, i.e., to detect the test in the presence of a mask stimulus. In these experiments, the only cue which differentiates the stimuli is their contrast (the mask-plus-test composite will have a higher contrast than the mask). Two stimulus combinations were used: both mask and test were Gabor patches or both were full-sized gratings.

The mask and test stimuli were presented on alternate frames of the display and, since the display had a frame rate of 120 Hz, the alternation of mask and test led to perceptual fusion at an effective frame rate of 60 Hz for each. The advantage of using alternate frames was that the mask and test stimuli would have separate look-up tables, giving both stimuli the maximum grey-level resolution. A low-contrast test stimulus could still be specified to 8 bits of high precision, even in the presence of a high-contrast mask. When the mask stimulus was presented alone without an added test stimulus, it was alternated with a blank screen at mean luminance. This frame alternation resulted in a halving of the effective contrast of the stimuli, so that the actual contrasts presented were half of the values reported in the text and figures.

2.1.2. Observers

Observers were GT and MC, both females in their mid-twenties with normal or corrected to normal vision. GT was naïve to the purpose of the experiments, whereas MC is one of the authors.

2.1.3. Protocol

The discrimination experiments used a modified *two alternative forced choice* (2AFC) procedure. Stimuli were presented sequentially in groups of *three* presentations (a *trial*), with a stimulus presentation time of 100 ms and an inter-stimulus interval of 120 ms, where stimulus onset and offset had a square-wave temporal envelope. The observer knew that the second stimulus of the trial was always the mask stimulus alone. The mask–test composite could be either the first or the third stimulus presented in the trial. The remaining interval would present the mask alone. The observer had to decide which interval (1 or 3) differed from the reference (interval 2) and make an appropriate response on the computer keyboard, receiving aural feedback. This three interval procedure was developed for use in other experiments, not reported here, in which mask and test have different spatial waveforms and the observer may confuse the appearance of the mask and mask–test composite (e.g., Tolhurst & Barfield, 1978). This is not a problem in the basic contrast discrimination experiment presented here, but the use of three intervals was retained, since it makes no difference to the psychophysical measurement.

The response to each trial was logged directly on the PC. An experimental session would measure thresholds for eight different values of mask contrast, spanning the range from 0% to 100% in 10 dB steps, using eight interleaved staircases. All the stimuli in an experimental session would have the same geometry (e.g., all might be Gabor tests on Gabor masks). Trials for the different mask contrasts were presented at random until each had been presented five times. Then, the contrast of the test stimulus was adjusted for each of the eight staircases for the next block of five trials. If the observer correctly chose the interval containing the test five times out of five, the test contrast was reduced for the next trials by 1–3 dB (where step-size decreased as the experiment proceeded). If the observer responded correctly four times out of five, the test contrast was considered to be at threshold and left unaltered; otherwise the contrast was increased by 1–3 dB. A complete experiment consisted of 20 blocks (100 trials) for each mask contrast.

A second experimental session was run in which, apart from the extreme values of 0% and 100% contrast, the masking contrasts were different, being interleaved between the initial 10 dB steps.

At the end of the experiment, for each masking condition, the test contrast versus percentage correct data (i.e., the psychometric functions) were fitted by an *error function* (i.e., a curve of the form erf((contrast_dB − threshold_dB)/slope), constrained to asymptotes of 50% and 98% correct), using a *maximum likelihood* algorithm. The slope and position (threshold parameter) of the function were adjusted to give best fit to the experimental data. The discrimination threshold was taken by interpolation as the 74% correct point of the fitted function. Results presented are the means of the threshold, estimated over 100 or 200 trials, of both of the observers averaged together.

2.2. Modelling methods

Our Bayesian model of contrast coding by populations of V1 neurons has been described in detail before (Chirimuuta et al., 2003; Clatworthy et al., 2003). It has three stages:

1. *Simulating noisy neuronal responses*. This stage uses the Naka–Rushton equation, with multiplicative Poisson noise (but variance *twice* the mean), to simulate each V1 neuron's contrast–response function. Parameters of the Naka–Rushton curves take values measured electrophysiologically in macaque V1.
2. *Building the contingency table*. By simulating the noisy contrast–responses for a large number of trials, one can estimate the distribution $P(c|r)$, the probability of any particular contrast having been presented, given the particular response of a neuron or a group of neurons.
3. *Simulating the discrimination experiment*. For a large number of trials, the model uses this contingency table to estimate, from the noisy neuronal responses on each trial, the contrast presented in each of a pair of stimulus intervals. The model adjusts the difference between the pair of contrasts until it correctly identifies the higher contrast interval on 75% of the trials.

2.2.1. Simulating noisy neuronal responses

The Naka–Rushton (Eq. (5), Naka & Rushton, 1966) has been used to fit the contrast–response data of neurons in cat and monkey V1 (Albrecht & Hamilton, 1982; Gardner, Anzai, Ohzawa, & Freeman, 1999; Li, Peterson, & Freeman, 2003; Sclar, Maunsell, & Lennie, 1990; Tolhurst & Heeger, 1997).

$$R = R_{\max}\frac{c^q}{c_{50}^q + c^q}, \tag{5}$$

where R is *mean* neuronal response in spikes per stimulus trial, averaged across many trials. $R_{\max}$ is the maximum *mean* response. c is stimulus contrast (as defined by Eq. (2) above). q is an exponent which determines the steepness of the curve and it takes values around 2 on average for real V1 neurons (Albrecht & Hamilton, 1982; Sclar et al., 1990). c_{50} is the semi-saturation contrast (the contrast at which the neuron attains half its maximum spike rate) and this determines the position of the response curve along the x-axis (the contrast axis).

The Naka–Rushton function is elegant, but we shall show that it can give rise to predictions that may not always be credible, because response (or the probability of a response) never reaches zero until the contrast is zero. In some models, therefore, we have introduced a "hard threshold" (compare Barlow, Kaushal, Hawken, & Parker, 1987): response was set to zero if R turned out to be less than 1–2% of $R_{\max}$, while, at high contrasts, R was not allowed to rise above 96–99% of $R_{\max}$. The hard threshold was implemented by subtracting 1–2% from each calculated R value, and then setting all negative values to zero. These abrupt high- and low-contrast cut-offs produced a function with very nearly the same shape as the standard Naka–Rushton, but without the smooth S curves at the top and bottom. Simulations were run with and without the hard threshold. Tolhurst and Heeger (1997) showed that it may be almost impossible to determine from the V1 data if the neurons have a hard threshold, since the Naka–Rushton function is an excellent fit to the empirical contrast–response curves, even at very low response levels, whether or not they really have the hard threshold. Therefore, our model offers a means of testing the adequacy of the standard Naka–Rushton: if the psychophysical data are better fit by the model which uses the non-standard Naka–Rushton (i.e., with a hard threshold), there is a reason to believe that the standard equation may be empirically inadequate.

Equation (5) gives a deterministic relationship between contrast and mean response, R, for a model neuron. However, the responses of V1 neurons are subject to considerable variability or "noise". It has been found that the variance of response increases with increasing mean response level; i.e., the noise is multiplicative (Dean, 1981; Geisler & Albrecht, 1997; Itti et al., 2000; Tolhurst et al., 1981, 1983; Vogels et al., 1989; Wiener et al., 2001). The response distribution is similar to a Poisson distribution, except that variance is roughly *twice* the mean response, within the range found by Vogels et al. (1989) and Geisler and Albrecht (1997) in the monkey visual cortex. In the Poisson distribution (Eq. (6)), variance is equal to the mean:

$$P(x) = \frac{e^{-\mu}\cdot\mu^x}{x!}, \tag{6}$$

where x is an integer and μ is the mean response for the contrast in question, calculated according to Eq. (5).

Variance twice the mean is generated by a second Poisson random process (Eq. (7)) which uses the first noisy response, x, as its mean. This gives the probability $P(r|c)$ of the final response, r, of the neuron on one trial, given a particular contrast, c:

$$P(r|c) = \sum_x \left(P(x) \frac{e^{-x} x^r}{r!} \right) \tag{7}$$

r is an integer, an integral number of "spikes" per stimulus presentation. Fig. 2 is an example of a model contrast–response function, showing the mean response and square root of response variance. The curve is sigmoidal and relatively steep since the exponent, q, is at two. c_{50} is 0.1, the most usual value in monkey V1 as shown in Fig. 2B, the distribution of values recorded from macaque monkey V1 by Ringach (personal communication; see Ringach, Hawken, & Shapley, 1997 for methods).

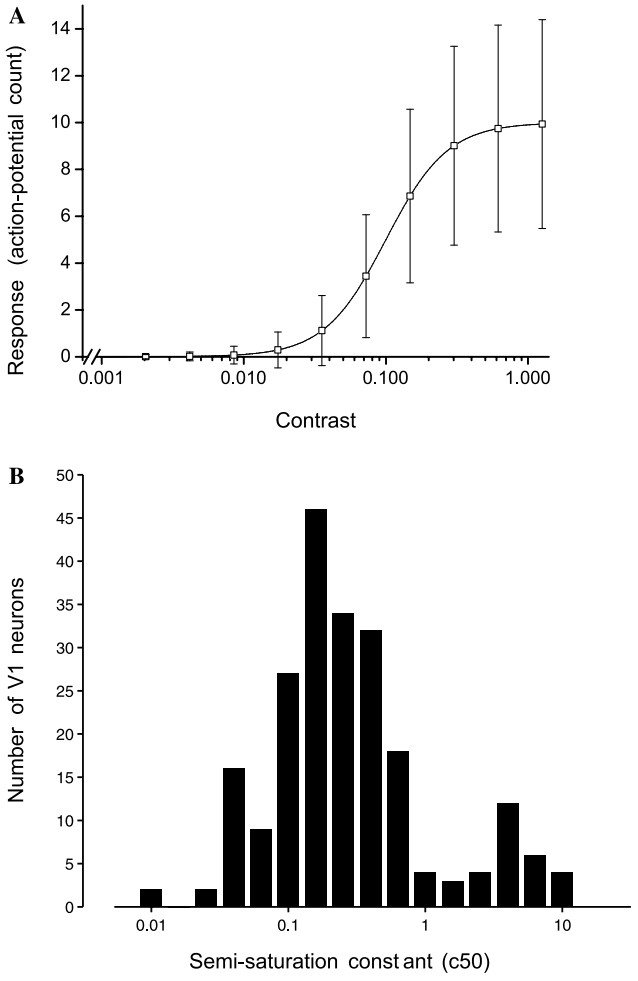

Fig. 2. (A) Typical contrast–response functions of the model. Model neuron has an R_{max} of 10, q of 2 and c_{50} of 0.1; there was no response threshold. Continuous line is the mean of Naka–Rushton equation; error bars show the square root of the response variance. (B) Histogram of c_{50} values of 219 neurons in macaque V1 (D.L. Ringach, personal communication).

2.2.2. Building the contingency table

To model contrast discrimination, it is necessary to know $P(c|\mathbf{r})$, the probability that a given contrast, c, has been presented given the responses, $\mathbf{r}$, of a set of n neurons. The contingency table between contrast and response is built up by repeatedly simulating noisy neuronal responses, and is the a posteriori probability distribution of an ensemble of neurons that allows the model to estimate the value of a stimulus contrast from a noisy neuronal response. $P(r|c)$ is calculated from Eqs. (5)–(7) for each of a neuron's possible instantaneous responses (taken to be from zero to four times R_{max}), for each of 311 contrasts from 0.001 to 1.26 in equal logarithmic steps of 0.01 log units (i.e., extending slightly past the physical contrast range for a pure sinusoid, though contrast >1 can exist e.g., as the fundamental of a square wave). The likelihood distribution can be summed across contrasts to give $P(r)$, which is the divisor of the Bayes equation (Eq. (1)):

$$P(r) = \sum_c P(r|c). \tag{8}$$

Bayesian statistics provide a framework for making inferences from noisy signals. If the relationship between contrast and response is non-deterministic, a good strategy for inferring the contrast of the stimulus that elicited a response is to infer that the contrast presented was that contrast for which the probability of its occurrence, given the response, was maximum. This is the *maximum* a posteriori rule: the probability of a contrast having occurred, given a particular response, $P(c|r)$ is calculated from $P(r|c)$ and $P(r)$ according to Bayes theorem (Eq. (1) in Section 1; see e.g., Mamassian, Landy, & Maloney, 2002 and Lee & Mumford, 2003; for use of this rule in other vision models). Since the simulations involve a *set* of model neurons, each with slightly different contrast–response function, their individual contingency tables must be combined so that one knows the overall *maximum* a posteriori, $P(c|\mathbf{r})$, where $\mathbf{r}$ is the set of responses of the different neurons. The individual likelihood distributions for the different model neurons are multiplied together, and divided by the joint $P(\mathbf{r})$ normalising term:

$$P(c|\mathbf{r}) = \frac{\prod_n P(r_i|c)}{P(\mathbf{r})} = \frac{\prod_n P(r_i|c)}{\prod_n P(r_i)}, \tag{9}$$

where r_i is the set of n individual responses, and $P(r_i|c)$ is the likelihood distribution of the ith neuron. This equation does not have an explicit prior term; instead, $P(c)$ is taken to be a flat distribution which makes the prior term irrelevant, so that the posterior distribution is equivalent to the likelihood distribution. Chirimuuta et al. (2003) found that a prior derived from the frequency of occurrence of different contrasts in natural scenes made little difference to their measure of information transmission in this model.

An alternative to the pooling rule of Eq. (9) would be simply to sum the action potentials of individual model neurons on each trial and not estimate individual likelihood distributions (Boynton et al., 1999; Heeger et al., 2000; Shadlen, Britten, Newsome, & Movshon, 1996). Such a rule ignores information conveyed by the differential firing rates of the neurons. The effects of the different types of pooling rules are discussed by Clatworthy et al. (2003) in the context of a Bayesian model of contrast identification. Preliminary simulations of the discrimination experiment showed that this alternative pooling rule made no qualitative difference to the model predictions, its use being equivalent to a decrease in neuronal firing rate, and so was not investigated further.

2.2.3. Simulating the contrast discrimination experiment

To simulate a discrimination experiment, one must first choose the number (n) and the parameters of the model neurons that will perform the task. In some simulations, only one model neuron was used, and the effects of changing its parameters (R_{max}, c_{50}, q and hard threshold) were investigated.

In other simulations, a *set* of neurons was used, the choice of parameters giving three different types of model for fitting the psychophysical data. Two of these models included data from a set of Naka–Rushton fits of 219 V1 neurons provided by Ringach (personal communication). The data collection Methodology is given by Ringach et al. (1997). In the first model, the model neurons each have different c_{50} values randomly selected from the distribution of monkey c_{50}'s (see Fig. 2B) to give a representative sample of these data. R_{max} varies in order to fit the data but stays the same for all neurons in the set, while q is fixed at 2 for all model neurons. Note that when $q = 2$, R is a function of contrast energy (Heeger, 1992b), and the equation with an exponent of 2 has been shown to have theoretical interest, as it maximises the information transmission rate of the neuron, given limited coding resources (Gottschalk, 2002).

In the second type of model, c_{50} is again sampled from the distribution of c_{50}'s recorded in the monkey, but now q and R_{max} can also vary amongst the neurons in the set. The motivation for this model was that we found that the monkey data provided by D.L. Ringach showed some correlation between c_{50} and q (correlation coefficient $= -0.177$), such that the cluster of neurons with high c_{50}'s (see Fig. 2B) tended to have a low q. If this were not the case, these neurons would never show any response because, with a c_{50} beyond the physical range of contrasts, the contrast–response function's curve must be shallow if the neuron is to begin responding to any stimuli. Similarly, if the high c_{50} neuron is to give a sensible response to high contrast that is comparable with the other neurons, its R_{max} parameter must be higher, because it will only respond at a fraction of this hypothetical R_{max} to any real stimulus. In the light of the correlation observed in the monkey data, therefore, we related q for the set of model neurons to c_{50} in the following way:

$$q = \frac{3 - \log 10(c_{50})}{2}. \tag{10}$$

such that $q = 2$ for the lowest c_{50} neurons falling to $q = 1$ for the highest c_{50} neurons in the set, although some real neurons had values outside this range. Likewise, the modelled R_{max} was fixed so that all neurons would give the same mean response at 100% contrast (R_{100}):

$$R_{max} = R_{100}(c_{50}^{q} + 1) \tag{11}$$

The third type of model allowed R_{max} and c_{50} to be absolutely free parameters. Thus, adherence to the true V1 distribution of c_{50} was sacrificed, and the set of c_{50}'s was no longer sampled from the empirical distribution, but was selected arbitrarily to give the best possible fit to the psychophysical data. R_{max} also varied freely for the different neurons of the set, while q was fixed at 2 for all neurons, and the threshold parameter could be varied to improve the fit, but was fixed at the same value for all neurons in the set.

The parameters of the models were adjusted by trial and error to give the best fit to the experimental data of the basic contrast discrimination experiments. The number of neurons (n) would be fixed, since R_{max} and n trade off (see Section 3); choice of n included the choice of which c_{50} values were used. Then, values of R_{max} and q, for instance, could be sought by fairly systematic coverage of the parameter space. An automatic optimisation algorithm was not used, since each iteration of the procedure would require complete recalculation of the contingency tables and simulation of complete experimental staircases at 14 pedestal contrasts; the variability in the outputs of model staircases at each iteration of the procedure introduced discontinuities in the function relating changes in the parameters to fit to the psychophysical data, incompatible with the use of these algorithms. Once the neuronal parameters had been chosen, the $P(c|\mathbf{r})$ contingency table for that neuron or set of neurons was calculated as above.

The present model makes no explicit reference to stimulus size: the only modelled parameter of the Gabor patch or full-sized grating is their contrast. However, if the experimental results for two different stimulus configurations needed a different number of neurons in the best-fitting model, one could argue whether an increase in the number of model neurons might reflect an increase in the stimulus size (see Section 4).

With a given model and response–contrast contingency table, the simulation took a set of the 14 mask contrast values that were used in the psychophysical

experiment and a corresponding set of test contrast values at *estimated* threshold levels. These contrasts were halved, just as the psychophysical contrasts were effectively halved by being displayed on alternating frames. The model generated 10,000 pairs of noisy responses to each mask and mask–test pair, using Eqs. (5)–(7) to calculate the response to a stimulus of the masking contrast and to a stimulus whose contrast was the sum of the mask and the test. The contrast of each of the paired stimuli was inferred from the contingency table. A model staircase adjusted the test contrast until the higher contrast in the pair was correctly inferred on 75% ± 0.6% of a set of 10,000 trials. The goodness of a given model's fit to a set of experimental results was quantified as the *mean-squared-error* (mse) between the 14 data points and model predictions:

$$\text{mse} = \frac{\sum_{i=1}^{14}(y_i - \widehat{y}_i)^2}{14}. \quad (12)$$

2.2.4. Modelling the psychometric function

We also used the Bayesian model to simulate the psychometric function for contrast detection and discrimination. The models simply calculated the percentage correct identification of the interval containing the non-zero contrast (detection) or the contrast-increment (discrimination) over 10,000 trials, as a function of test contrast. The modelled psychometric functions were then fitted with the Weibull function (Robson & Graham, 1981):

$$\text{Percentage correct} = 100\left(1 - 0.5\exp\left[-\left(\frac{c}{\alpha}\right)^{\beta}\right]\right), \quad (13)$$

where c is the contrast of the test stimulus in the detection experiment, α is the contrast at which the observer correctly identifies the 2AFC interval on 81.6% of trials, and β indicates the slope of the function. The main purpose of the simulations was to see how changing model parameters affected the steepness of the slope (β), which is about 3–4 in psychophysical detection experiments (e.g., Robson & Graham, 1981) but is less than 2 for single V1 neurons (Tolhurst et al., 1983).

3. Results

3.1. Psychophysical results

The psychophysical results confirmed the well-known dipper effect for both observers, and for both stimulus conditions (Gabor mask and test, and grating mask and test). The data are plotted in Figs. 7–9, along with the model predictions, and these results will be discussed further in comparison with the model results (see below).

3.2. Single-neuron simulations

Single-neuron simulations were carried out in order to see clearly the effects of changing the different parameters of the Naka–Rushton (Eq. (5)) on contrast discrimination. Fig. 3 presents the results of these simulations, plotting test threshold against mask contrast. The model curve is not smooth because model thresholds were only ever simulated at the 14 mask contrast values used in the psychophysical experiment, in 5 dB steps apart along the x-axis. The first thing to notice is that this model does generate a dipper function, with all model predictions showing facilitation and masking, except for the case when the Naka–Rushton exponent, q, takes the low value of 1.

Fig. 3A shows the effects of changing $R_{\max}$, the maximum mean firing rate, where firing rate is taken to be the number of action potentials per stimulus interval. The figure shows that *increasing* $R_{\max}$ steadily *decreases* the discrimination thresholds; that is, the model performs better if its single neuron is able to produce more action potentials per interval. The shape of the dipper does not change much, except that the dip gets deeper as $R_{\max}$ increases (though it is still not particularly deep in comparison with the psychophysical findings of Figs. 7–9). This improved performance is to be expected in the light of the finding of Clatworthy et al. (2003) that *accuracy*, their measure of the model's performance in an identification task, was also improved by increasing $R_{\max}$. We have previously argued that typical V1 values of $R_{\max}$ are likely to be at the lower end of the range shown in Fig. 3A (Clatworthy et al., 2003).

In Fig. 3B, $R_{\max}$ is fixed at 10 while c_{50} takes four values from 0.01 to 0.60, within the range of values recorded in monkey V1. c_{50}, the contrast at which the neuron achieves half its maximum firing rate, is the parameter that determines the sensitivity of the model neuron to low contrasts and the value of the higher contrast to which it will begin to give a saturating response, in other words, determining the contrast range over which the neuron gives a differential response. Increasing c_{50} shifts the model discrimination function upwards and rightwards: only the neuron with the lowest c_{50} is sensitive to the smallest contrast differences but it soon saturates, with threshold rising vertically at a lower contrast than for any of the other neurons, since the dynamic range is the same, whatever the value of c_{50}.

Fig. 3C compares pairs of neurons with the same $R_{\max}$ and c_{50}, with and without a 2% "hard threshold" in the response function. The plot shows two pairs of results, one pair has an $R_{\max}$ of 10, the other of 50. In both cases, the absolute detection thresholds of the neuron with the response threshold are the higher. But these neurons now show a considerably stronger facilitation effect so that their lowest discrimination thresholds drop well below those of the other neuron. At high contrasts

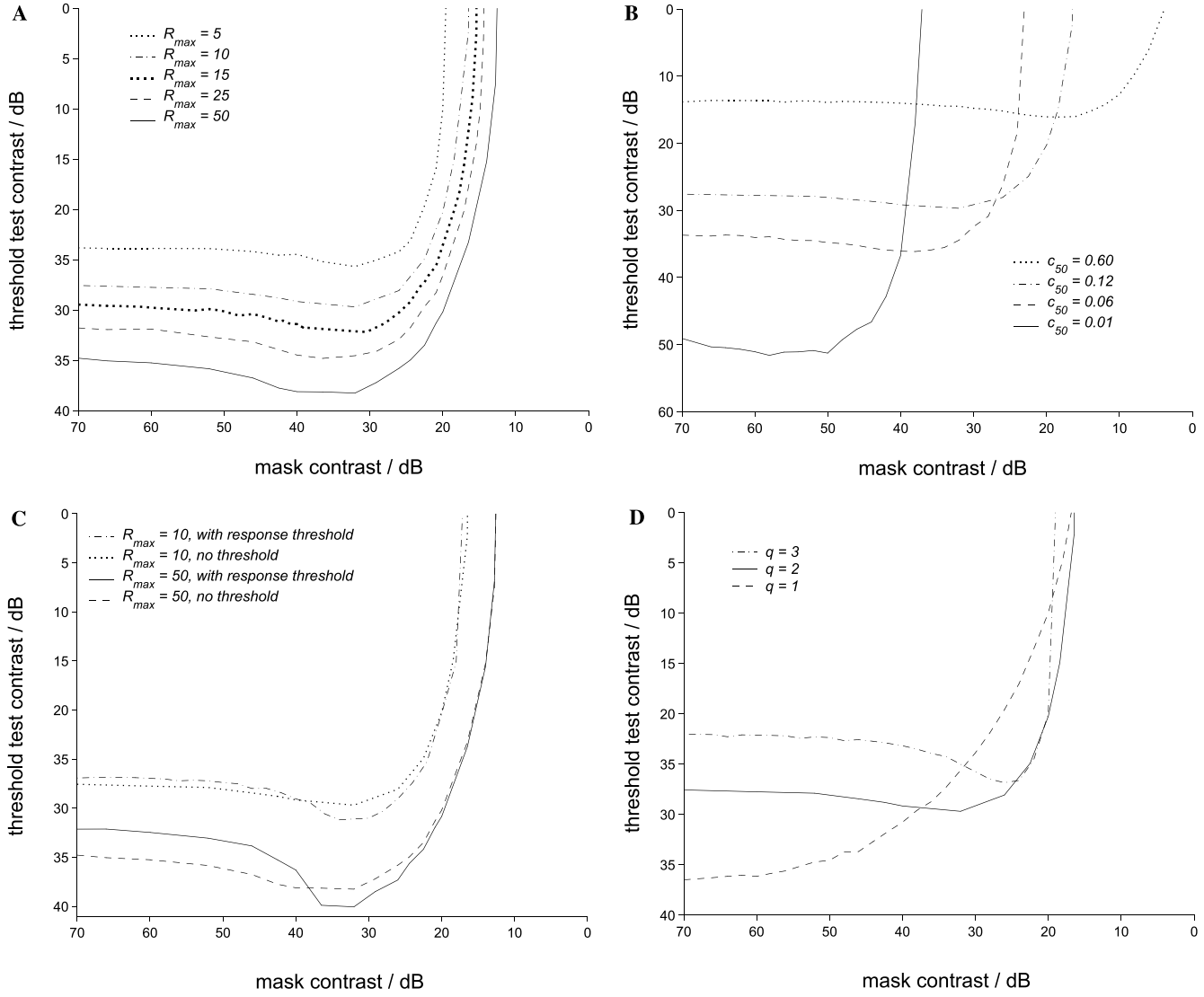

Fig. 3. Results of the single-neuron model, showing the effects of changing different contrast–response function parameters. Predicted threshold increment contrast is plotted against mask contrast. (A) R_{max} takes different values between 50 and 5; $q = 2$, $c_{50} = 0.121$, no response threshold. (B) c_{50} takes four different values between 0.01 and 0.60; $q = 2$, $R_{max} = 10$, no response threshold. (C) Comparing results with and without response threshold; $R_{max} = 10$ or 50, $q = 2$, $c_{50} = 0.121$. (D) q at 1, 2 or 3; $R_{max} = 10$, $c_{50} = 0.121$, no response threshold.

the results with and without response threshold are indistinguishable. The effect must occur because the neurons with a hard threshold will, for the set of stimulus contrast just under their detection thresholds, only need a tiny contrast increment to go from an unresponsive state to one of high responsivity, thus generating very low discrimination thresholds (Barlow et al., 1987; Tolhurst & Barfield, 1978).

As noted above, all of the single-neuron predictions showed facilitation except for the case in which $q = 1$. This result is shown in Fig. 3D. Nachmias and Sansbury (1974) first hypothesised that it is the acceleration of a channel's contrast–response function that is responsible for the facilitation effect (see Fig. 1, Section 1). When $q = 1$, the contrast–response function of the neuron does not accelerate, being nearly linear at low contrasts, and, as to be expected from the hypothesis, does not give a dipper. Increasing q from 2 to 3 (i.e., making the contrast–response function more strongly accelerating) gives an even deeper facilitation result. One may note that the strong facilitation of the psychophysical transducer function (Legge & Foley, 1980; see Fig. 1) is achieved by a steep exponent between 2 and 3.

Before discussing the multiple-neuron simulations, the characteristic shape of the single-neuron dipper should be noted. All of the dippers in Fig. 3 show an almost vertical rise at moderate to high contrasts. This is due to the limited dynamic range of a single neuron (Albrecht & Hamilton, 1982; Tolhurst et al., 1981, 1983). A single neuron with a steep contrast–response function (q around 2) only gives a differential response over a limited contrast range before saturation. It

therefore follows that the psychophysical results *cannot* be due to the responses of any such single neuron, since no psychophysical dippers have been reported to have this shape.

Although the single-neuron models do show facilitation and masking, the results do not resemble the psychophysical dipper (Fig. 7). Simulations of multiple neurons were therefore run in order to see if a *pool* of different neurons, responsive over different portions of the contrast range, would give more realistic-looking results.

3.2.1. Simulations of multiple neurons

Fig. 4 presents the results of including more neurons in the simulation *without* changing the values of c_{50}. Clatworthy et al. (2003) reported that increasing the number of neurons in their model had exactly the same effect as increasing $R_{\max}$, as long as the model neurons all had the same c_{50}. Performance would be identical if the total number of action potentials (i.e., $R_{\max}$ times the number of cells) were the same in both cases. The results here are in agreement with this finding. Simulations were run with sets of two, three and five neurons, with an $R_{\max}$ of 5 or 10. Fig. 4 compares these simulations with the corresponding single-neuron simulations with $R_{\max}$ at 10, 15, 25 and 50. The resulting dippers are very nearly identical.

Fig. 3B above showed that a low c_{50} neuron is only able to discriminate low contrasts, and a high c_{50} neuron is only able to discriminate high contrasts. From this, it is reasonable to conclude that the psychophysical dipper, which shows that humans can discriminate contrasts over a wide range, reflects the operation of a group of neurons with a range of different c_{50} values. In order to test this idea, simulations were run with groups of neurons with c_{50}'s sampled from the data of Ringach (Fig. 2B, personal communication; Ringach et al., 1997) measured in monkey V1.

As expected, simulation with the "realistic" c_{50} distribution produced dippers which no longer shot up vertically at high mask contrasts. This can be seen in Fig. 5 which plots the results of populations of 8, 16 and 36 neurons sampled from the monkey c_{50} distribution, with all neurons having the same q of 2, and $R_{\max}$ of 10, and a 2% response threshold. One sees a similar effect of changing the number of neurons as presented in Fig. 4 above, giving a set of almost parallel graphs, even though the extra neurons added to the model will have different interleaved c_{50} values. The model with the most neurons has the lowest thresholds and has the deepest dip. Notice also that this 36-cell model is also the one with the most gradual slope at high contrasts, because of the inclusion of more high c_{50} neurons.

Fig. 3C showed that a 2% hard response threshold produced a deeper dip than the single-neuron model without the threshold. Fig. 6 shows that the magnitude of the effect increases if the threshold is raised in the multiple-neuron model as well.

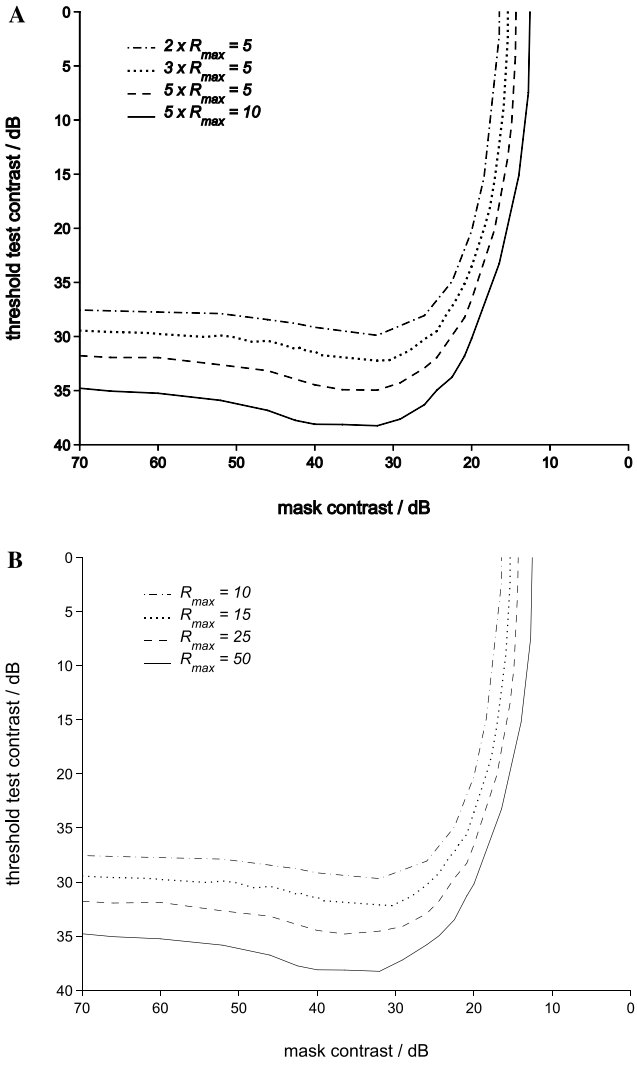

Fig. 4. Results of multiple-neuron model (A) compared with single-neuron model (B), showing the equivalence of the models if "total number of action potentials" (i.e., $R_{\max} \times$ number of neurons) is the same. In all models, $c_{50} = 0.121$, $q = 2$, and there was no response threshold.

3.2.2. Fitting the model to psychophysical data

This section assesses the fit of the model to the psychophysical data. Since there is such a clear trade-off between $R_{\max}$ and n (see Fig. 4), neither of these parameters can be fixed by comparison with experimental data: an appropriately low detection threshold, for example, could be achieved by increasing either $R_{\max}$ or n. In order to fit the data, the parameters of the model were adjusted by trial-and-error until a fit was achieved that was both reasonable looking and had the lowest mse (Eq. (12)).

The data to be fit (Fig. 7) were the thresholds of the two observers, MC and GT averaged together (error bars show standard error of the mean); the two different stimulus conditions were fit separately, i.e., small Gabor patch stimuli, or large grating stimuli. For all of the model types used, it was found that a good fit—in

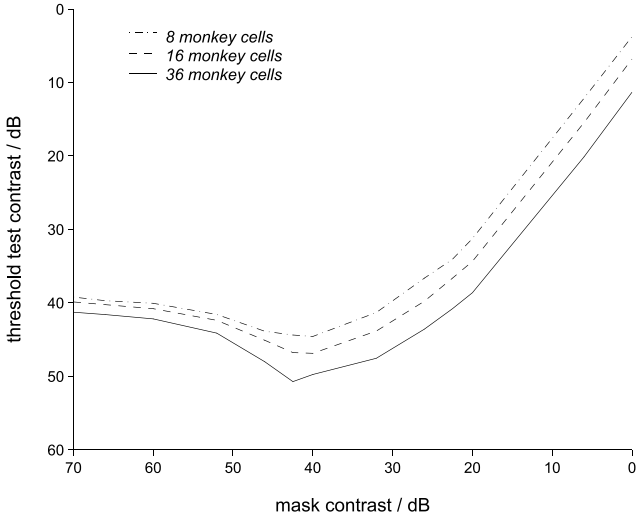

Fig. 5. Results of multiple-neuron "monkey model", for populations of 8, 16 and 36 cells. $R_{max} = 10$ and $q = 2$. c_{50} is sampled from the distribution of values measured by Ringach (personal communication). Model neurons have a 2% response threshold.

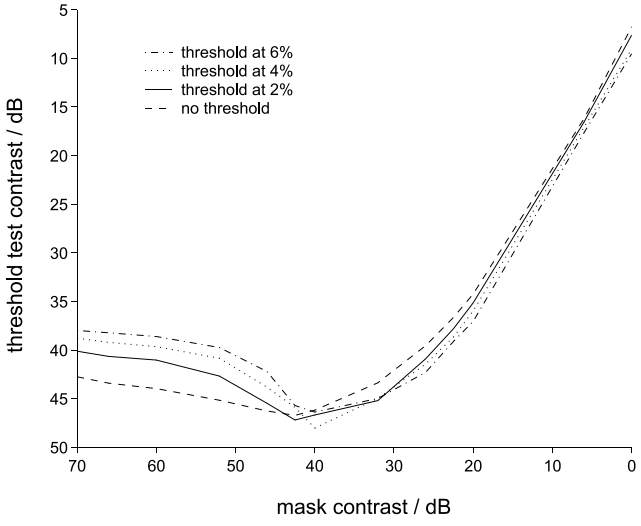

Fig. 6. The effects of changing the "height" of the response threshold in the monkey model. Threshold takes values of 0%, 2%, 4% and 6%. $R_{max} = 10$ and $q = 2$ and $n = 18$. c_{50} is sampled from the distribution of values measured by Ringach (personal communication).

particular, a dip that was deep enough—could only be achieved if the model neurons had a hard response threshold of 1% or 2%. The parameters to be adjusted depended on the particular model type. For the first model, a simple monkey model, c_{50} was sampled from the data set provided by Ringach (personal communication), while different values of n were tried. For each value of n, R_{max} was varied freely, although it was the same for all of the model neurons in the set, and q was fixed at 2. Fig. 7 shows this model's fits to the Gabor patch (A) and grating (B) data.

The better fit is achieved for the experiment in which both mask and test are Gabor patches (part (A), mse = 16.84) as opposed to the gratings experiment (part (B), mse = 79.93). However, in both cases the model shows a systematic error whereby the fits are good for the lowest contrast masks, up until the minimum of the dip. At that point, the psychophysical data rise sharply until the few highest contrasts, where the slope of the graph flattens off. The model dipper, on the other hand, rises gradually out of the dip, and has a steeper slope at the highest contrasts, the opposite pattern to the psychophysical dipper. The pattern of

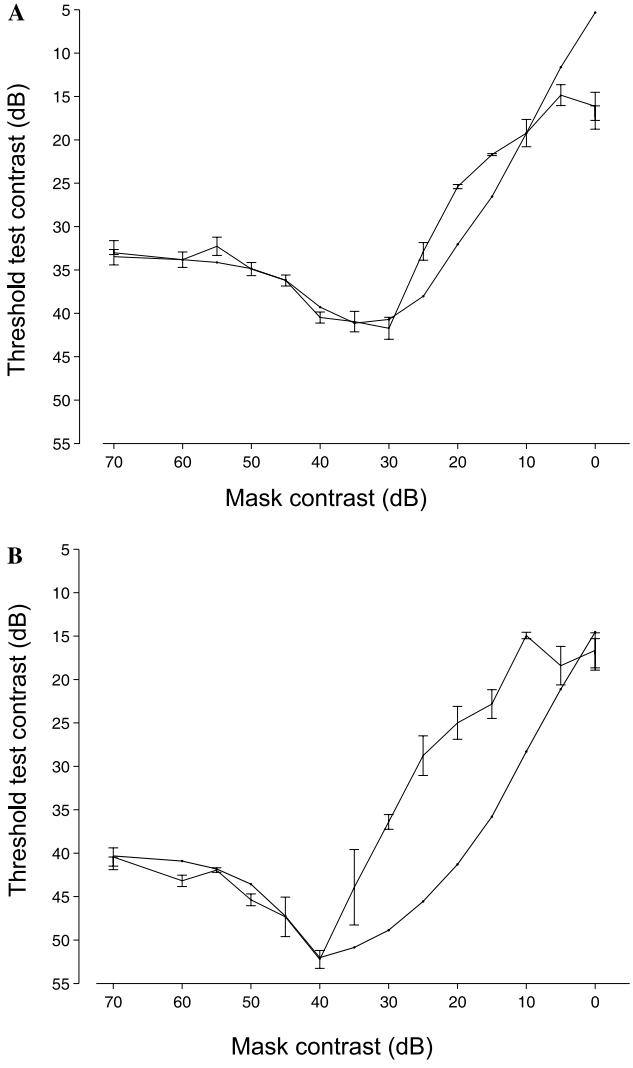

Fig. 7. Mean experimental results of observers (± standard error of the mean) and best-fitting basic monkey model predictions (small markers). Experimental stimuli were 2.67 c deg^{-1} vertical gratings or Gabor patches. 0% contrast mask is plotted as 70 dB. (A) Both mask and test were Gabor patches. Mean of two observers, GT and MC. Best-fitting monkey model parameters: $n = 12$; $R_{max} = 8$; $q = 2$; threshold = 2%; mse = 16.84. Total action potential count at 100% contrast = 80. (B) Both mask and test were gratings. Mean of two observers, GT and MC. Best-fitting monkey model parameters: $n = 22$; $R_{max} = 30$; $q = 2$; threshold = 1%; mse = 79.93. Total action potential count at 100% contrast = 549. Threshold test contrast is plotted against mask contrast in dB. Note that these contrast values are double the values actually presented to the observer or to the model.

psychophysical results (the flattening at higher contrasts) can be discerned also in the experiments of Boynton et al. (1999), Bradley and Ohzawa (1986) Foley (1994), Itti et al. (2000), and Meese (2004). It should be noted that, although this simplistic model of V1 response pooling from a population of neurons has produced only a moderate fit, it is a far better than any fit that could be achieved with almost all single-neuron models (Figs. 3 and 4).

In order to try to address this obvious discrepancy, less simplistic monkey models were developed, in which, again, c_{50} is sampled from the neurophysiological data set, but R_{max} and q are not constrained to be the same for all neurons. Instead, they correlate with the c_{50} of the model neurons in the set so that a high c_{50} neuron will have a low q and high R_{max} so that its contribution to the model will be boosted (Eqs. (10) and (11)). It was thought that the increased effectiveness of the high c_{50} neurons would give the model dipper a shallower slope at high contrasts, more like the psychophysical one. The correlation between q and c_{50} was observed in the data set, and the correlation between c_{50}, q and R_{max} is necessary if these neurons are to have any significant response within the physical contrast range (see Fig. 2B).

R_{100}, the mean firing rate of all neurons at 100% contrast, n and the value of the threshold, were adjusted to give the best fits to the psychophysical data. Fig. 8 plots the elaborated model fits with the data, again for the mean of MC and GT's thresholds for the Gabor patch (A) and gratings (B) experiment. There is a slight improvement in the mse for both experiments (Gabor patches mse = 13.94 as opposed to 16.84 for the first model, gratings mse = 76.89 as opposed to 79.93), but the large systematic error is still apparent at masking contrasts higher than the minimum of the dip. The new model has improved the fit by bringing thresholds down for the highest mask contrasts. But the model dipper's slope is also now shallower at the few mask contrasts just after the dip, which is actually contrary to the pattern of the psychophysical data.

In the case of both of the models discussed so far, the high mse values for the gratings experiment, and the error in the Gabor experiment, appear to be due to the inability of the model to match the sharp rise in thresholds between the middle- and high-contrast range. Perhaps the fits could be improved by reducing the number of middle c_{50} neurons, while keeping those with the highest c_{50} so that thresholds level off at high contrasts. Note also that the model does not predict the high-contrast curvature of the psychophysical dippers, where the threshold values flatten off at the very highest mask contrasts. This pattern is not thought to be a peculiarity to these two observers, or due to an error in the experimental set-up (e.g., monitor luminance saturation at the highest contrasts) since, as mentioned above, it

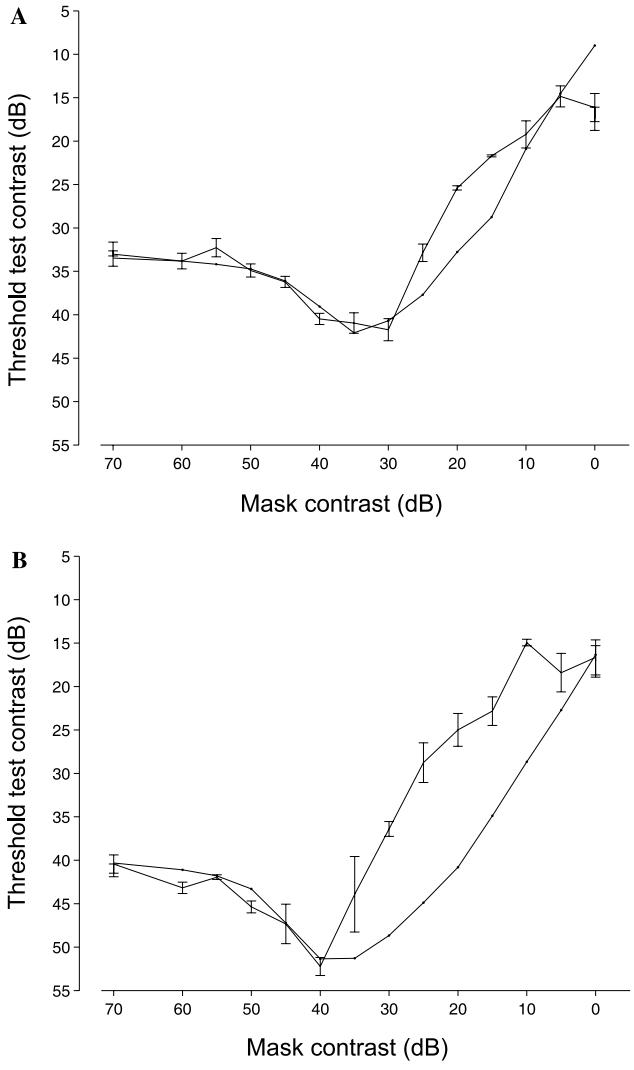

Fig. 8. Mean experimental results of observers (± standard error of the mean) and best-fitting elaborated monkey model predictions (small markers). Experimental stimuli were 2.67 c deg^{-1} vertical gratings or Gabor patches. 0% contrast mask is plotted as 70 dB. (A) Both mask and test were Gabor patches. Mean of two observers, GT and MC. Best-fitting monkey model parameters: $n = 12$; $R_{100} = 8$; $q = 2$; threshold = 2%; mse = 13.94. (B) Both mask and test were gratings. Mean of two observers, GT and MC. Best-fitting monkey model parameters: $n = 22$; $R_{100} = 25$; $q = 2$; threshold = 1%; mse = 76.89. Threshold test contrast is plotted against mask contrast in dB.

can also be seen in the results of Boynton et al. (1999), Bradley and Ohzawa (1986), Foley (1994), Itti et al. (2000), and Meese (2004); though these authors do not generally comment on it.

Thus, the third model no longer samples c_{50}'s from the physiological data set, but uses an arbitrary c_{50} distribution which need no longer peak at the mid-contrast range, and might even be bimodal. q is fixed at 2 as in the first model, and the threshold at 2%. R_{max}, however, is allowed to vary between the different neurons of the set, which is equivalent to changing the *number* of neurons with a particular c_{50} (see Fig. 4 above; in effect,

the number of neurons is still a variable parameter); this way of changing the shape of the c_{50} distribution was simply more convenient, and more efficient, than changing the number itself, though the choice of a small number of c_{50} values and larger R_{max} values may make the model dipper functions less smooth with masking contrast.

Fig. 9 plots the fits of the third model with the psychophysical data. For both experiment types, mse is very much reduced (Gabor experiment, mse = 2.53; gratings experiment, mse = 11.99), as one would expect from the effective increase in the number of degrees of freedom in the more arbitrary choices of c_{50} and R_{max}. There is no longer the systematic error of the model dipper greatly undershooting the psychophysical thresholds just after the minimum, and overshooting at the highest contrasts. This fit is achieved by using a "U-shaped" c_{50} distribution (see legend to Fig. 9) which, in a reversal of the physiological distribution, peaks at the highest and lowest values and has few or no neurons at $c_{50} = 0.1$, the peak of the physiological distribution! We will consider a possible justification for this apparently arbitrary choice of c_{50} values in Section 4.

Finally, note that there is some consistency between all three models in the fit parameters used for the gratings versus the Gabors experiments. In all cases, the product of n and R_{max} (or the sum of the R_{100} values for Fig. 9) is 4–6 times higher for the grating-on-grating results than for the Gabor-on-Gabor results (see legends to Figs. 7 and 9). These models cannot explicitly simulate the different sized stimuli used in the experiments (gratings versus Gabor test patches); instead, they fit the data-sets of the different experiments in exactly the same way, regardless of stimulus size used, using "contrast" as the only experimental variable. However, the best-fitting product of n and R_{max} parameter values can be interpreted as reflecting the effects of changing stimulus size. Depending on the model, the total number of spikes evoked by a stimulus of 100% contrast ranged from 47 to 549, the smaller values being for Gabor patches.

3.2.3. The model psychometric function

Detection experiments (i.e., mask contrast of zero) were simulated and the psychometric functions generated by the models were fitted with the Weibull function (Eq. (13)). The focus of interest was the β parameter, which describes the steepness of the function. Tolhurst et al. (1983) observed that, while psychophysical or behavioural psychometric functions are steep (having a β typically around 3–4 at detection), the response probability ("neurometric") functions of single V1 neurons are shallower (with a β around 2) and so cannot, alone, account for the psychophysical result. Tolhurst et al. proposed "probability multiplication", whereby detection would be possible only if, say, four neurons were active together.

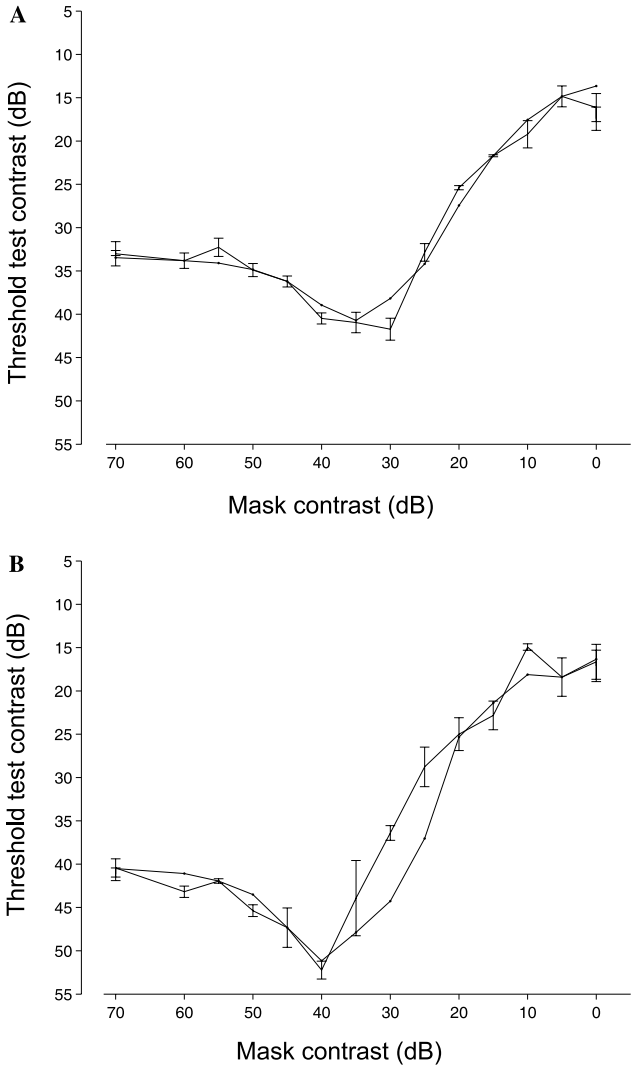

Fig. 9. Mean experimental results of observers (± standard error of the mean) and best-fitting model predictions, using optimal distribution of c_{50} and R_{max}. Experimental stimuli were 2.67 c deg^{-1} vertical gratings or Gabor patches. 0% contrast mask is plotted as 70 dB. (A) Both mask and test were Gabor patches. Mean of two observers, GT and MC. Best-fitting model parameters: $n = 8$; c_{50}= {0.08, 0.09, 0.1, 0.5, 0.7, 1, 3, 6}; R_{max} = {12, 6, 12, 6, 6, 12, 20, 90}; $q = 2$; threshold = 2%; mse = 2.53. Total action potential count at 100% contrast = 47. (B) Both mask and test were gratings. Mean of two observers, GT and MC. Best-fitting model parameters: $n = 5$; c_{50}= {0.05, 0.07, 1, 3, 6}; R_{max} = {60, 24, 30, 90, 90}; $q = 2$; threshold = 2%; mse = 11.99. Total action potential count at 100% contrast = 115.

β was determined for the psychometric functions of single model neurons, or for populations of up to 23 neurons, with c_{50} sampled from the distribution of V1 values (Fig. 2b), with an R_{max} of 10 and q of 2, with and without a 2% response threshold (i.e., similar to the models of Fig. 7). It was found that the populations *without* threshold always had a shallow psychometric function with a β under 2 (1.75–1.99, increasing slightly with number of neurons). If the Naka–Rushton had an added hard response threshold, however, the single

neuron had a steeper psychometric function ($\beta = 2.25$), and β rose steadily up to 4.20 for the populations of neurons. Similarly, the β of a single neuron was found to increase with R_{max} only if the neuron had the hard threshold. With such a threshold, increasing R_{max} from 5 to 50 caused β to increase from 2.2 to 3.5, whereas β remained under 2 if the neuron had no threshold. Thus, the steepness of the psychophysical psychometric function for contrast detection can only be simulated if the model neurons have a hard threshold, consistent with the finding that only simulations which include the response threshold give a dipper deep enough to match the psychophysical data.

We also simulated psychometric functions for contrast discrimination, plotting percentage correct in the simulations against the magnitude of the contrast increment. With parameters like those used to fit Figs. 9A and 7A, we found β values of 3.36 and 2.86 for detection, respectively; both models predicted lower values of β for contrast discrimination experiments. For instance, predicted β fell to a minimum of 0.97 (Fig. 9A settings) and to 1.07 (Fig. 7A settings) at a mask contrast of 0.1, just higher than the dip, and β remained close to 1 for all of the higher masking contrasts. In sum, the predictions of both the physiological monkey model and the non-physiological model are compatible with the experimental psychophysical findings of Bird, Henning, and Wichmann (2002), who showed shallower psychometric functions for contrast discrimination compared to detection.

4. Discussion

The results presented in this paper have shown that models of V1 neuronal responses which use Bayesian inference to discriminate two contrast intervals can, indeed, account for the dipping shape of the psychophysical result. However, the models which used a straightforward physiological distribution of the sensitivity parameter, c_{50}, produced a systematic error in the fits to the psychophysical data, in that the slope of the dipper at highest contrasts was predicted to be much steeper than for any of the experimental data sets. A model which used an apparently non-physiological c_{50} distribution gave a more satisfactory fit to the data. It is interesting to ask what can be concluded from these findings about the neural basis of contrast discrimination, given the number of assumptions and simplifications made in the development of the model.

We assume that it is processing in V1 which is exclusively responsible for setting the psychophysical thresholds, and that macaque V1 is a good model of human primary visual cortex. The latter assumption is supported by reports of dipper functions in a behavioural study in monkeys, very similar to the human ones (Kiper & Kiorpes, 1994), and by studies of contrast sensitivity in macaques and humans (De Valois, Morgan, & Snodderly, 1974). The failure of the models using the overall c_{50} distribution from V1, in comparison to the model using a non-physiological distribution, may be evidence against the former assumption, and may instead suggest that some extrastriate area, which has a different distribution of neuronal contrast sensitivities, is a key area for contrast discrimination, at least at high contrasts. However, there is some good evidence for this assumption, coming from the fMRI studies of Boynton et al. (1999) and Ress and Heeger (2003) which showed that the BOLD signal in early visual areas seems to predict psychophysical performance in a contrast discrimination task and a detection task, respectively. Still, the nature of the relationship between the BOLD signal and neural activity is not known, and this bears on the question of how one should model the activity of populations of neurons. Heeger et al. (2000) analysed a large body of human fMRI data and monkey electrophysiological data on contrast–responses in V1, and showed that the fMRI responses were proportional to the *average* neuronal firing rates at each contrast. Boynton et al.'s (1999) model for predicting dipper functions from fMRI data used the average BOLD signal across the responding voxels. We shall return to discussion of this model later. On the other hand, the Bayesian pooling rule used in our models incorporates information about the differential firing rates of individual neurons which is lost in a simple average, but we have assumed that the only information available for decision making is the integral number of action potentials generated by a neuron during a stimulus presentation; it may be that the temporal pattern of activity during a stimulus presentation might convey extra information (e.g., Reich, Mechler, & Victor, 2001).

For the models which took c_{50} values from the distribution recorded in monkey visual cortex, the best fits to the psychophysical data were achieved by small sets of 12–22 model neurons, with R_{max} ranging from 8 to 30. A single-neuron model could not fit the data because one neuron alone, with its limited dynamic range, was not able to give low enough discrimination thresholds over the entire contrast range. However, it *cannot* be concluded that, for example, only 12–22 neurons in the observer's brain were contributing to psychophysical performance, and that they were firing at a rate of R_{max} since the effects of increasing n and increasing R_{max} are equivalent in this model. Furthermore, the estimates of n could be too small if there were significant correlations in the response noise of different neurons (see Shadlen et al., 1996).

A trend shown by all three of the models was that the best-fitting product of n and R_{max} for the experiment in which both mask and test were large gratings was higher than for the experiments using Gabor patches. The

larger product was required to bring down the threshold predictions, especially at low contrasts and may reflect the fact that a large stimulus will stimulate a greater number of neurons, with spatial probability summation giving increased sensitivity (Robson & Graham, 1981).

An important finding was that an acceptable fit, with model dippers as deep as the experimental ones, could only be achieved if the model neurons had a hard response threshold. This concurs with the argument of Barlow et al. (1987) that the dipper effect can be explained as the result of response thresholds. Experimentally-measured contrast–response functions of real neurons are fit equally well by smoothly accelerating Naka–Rushton curves and those with a hard threshold (Tolhurst & Heeger, 1997), and so one cannot decide between the two descriptions simply by looking at contrast–response functions measured in extracellular recording experiments. To discount or prove a hard threshold would require presentation of thousands of repetitions of low-contrast stimuli in an extremely stable recording situation. Nevertheless, there is some physiological evidence, from the intracellular recordings of Carandini and Ferster (2000), that neurons in cat primary visual cortex *do* have a hard threshold, and these authors argue that its purpose is to sharpen the neurons' orientation tuning. The hard threshold was also needed for the model to produce psychometric functions for contrast detection that are as steep as human psychophysical ones. The Naka–Rushton equation is very convenient, and it results from elegant modelling (Gottschalk, 2002; Heeger, 1992a); however, in some minute details (that may be difficult to measure) it may be crucially inadequate.

The first model (the simple monkey model) took a set of model neurons with c_{50}'s varying according to a physiological distribution measured by Ringach (personal communication), and with $R_{\max}$ and q fixed at the same values for all neurons. Real neurons, however, vary quite considerably in the shapes and amplitudes of their response contrast curves (Albrecht & Hamilton, 1982; Geisler & Albrecht, 1997) and the data from which the c_{50} distribution was drawn showed some correlation between c_{50} and the steepness of the contrast–response function. Therefore, it is not surprising that the simplest model of monkey cortex should prove inadequate over some of the contrast range. Still, a second monkey model, which did include such a correlation, did not perform substantially better than the first.

The third model, which did not use the macaque V1 c_{50} distribution, did not produce the error of the first two, having a shallow enough slope at the highest contrasts, and a steeper rise out of the dip. This improved fit was achieved by *minimising* the number of neurons with intermediate c_{50} around 0.1, which is actually the most common value in the physiological distribution! Since, as mentioned above, the macaque V1 is often assumed to be a good animal model of human primary visual cortex, it needs to be considered whether or not the failure of the more physiological models is evidence against this assumption. In fact, although Boynton et al. (1999) do show an impressive fit of BOLD fMRI data to the supposed non-linear transducer which describes the contrast discrimination dipper, there are systematic errors in their fits too, most especially for the observer illustrated in their Fig. 3.

One thing to be noted is that the monkey distribution (see Fig. 2B) does have a cluster of neurons with $c_{50} > 1$. Our artificial c_{50} distribution of the third model has a relatively higher proportion of neurons with $c_{50} > 1$, and it is the responses of these model neurons which bring down the steepness of the dipper slope at the highest contrasts. It may be the case that this is consistent with the physiology, if it is the cluster of insensitive neurons, and not the more numerous neurons with $c_{50} \approx 0.1$, that is responsible for the computations underlying the discrimination of gratings in the visible contrast range.

A reason for thinking that this might be so is in considering the anatomical separation between "M" and "P" neurons which is well-recognised subcortically (Kaplan & Shapley, 1986; Sclar et al., 1990), but is more controversial in striate and extrastriate cortex. M-cells have high contrast sensitivity and high temporal-frequency sensitivity and are thought to be the major source of input for the dorsal stream for motion processing. P-cells, on the other hand, have low luminance contrast sensitivity but sensitivity to chromatic contrast, and are thought to be the source of input to the ventral stream for form processing (Livingstone & Hubel, 1988). Perhaps, the bimodality of c_{50} that we were obliged to impose on our model reflects the input from the two populations of LGN cells; the mean c_{50} of M-cells is 0.11 and for P-cells is 0.50 (Sclar et al., 1990). Perhaps, the unimodal distribution in V1 as a whole (e.g., Fig. 2b) results from the summing of many displaced bimodal distributions from different retinal eccentricities (1–6 deg, as measured by Ringach (personal communication), Ringach et al., 1997), different monkeys, or even for different optimal spatial frequencies. M- and P-cells might underlie "transient" and "sustained" channels described psychophysically (Kulikowski & Tolhurst, 1973; Tolhurst, 1973), and it is interesting that the rising phase of the contrast discrimination function is steeper with stimuli that favour the transient system (Boynton & Foley, 1999; Hess & Snowden, 1992; Lehky, 1985). This is consistent with our data which show an initial steep rise at low or medium masking contrasts where M-cells would predominate and a shallower portion at high masking contrasts where P-cells would predominate.

However, it has been suggested that the M- and P-cell streams converge considerably in V1 (Hawken, Parker, & Lund, 1988). A recent study (Vidyasagar, Kulikowski, Lipnicki, & Dreher, 2002) found that in all layers of V1

there are a number of neurons which do seem to have a convergence of M and P input, with others showing either M or P characteristics. It is not unfeasible that the discrimination task at high contrasts involving, as it does, the processing of the form of static gratings, is a task mediated by neurons in the cortex which have mostly P-cell input and therefore high c_{50}.

An alternative explanation for the bias towards high c_{50} neurons required to fit the psychophysical data is that, at high mask contrasts, adaptation (contrast gain control) reduces the sensitivity of neurons, effectively increasing their c_{50}. Contrast gain control or "normalisation" was studied in macaque V1 by Carandini, Heeger, and Movshon (1997), and it is thought that the compressive part of neurons' contrast–response functions is, in fact, due to the gain-control mechanism (Heeger, 1992a). By analogy with light adaptation, it is possible that contrast gain control may shift the dynamic range of a neuron to a new contrast range, allowing it to discriminate in that range. However, if one considers the time course of gain-control mechanisms, it does not seem likely that this can reconcile the neurophysiology to the c_{50} shift required to fit the psychophysical data. In the psychophysical experiment, stimulus intervals lasted just 100 ms. This is not necessarily less than the time in which it is possible for gain-control mechanisms to become effective, for Albrecht, Geisler, Frazor, and Crane (2002) report a fast gain-control mechanism in monkey V1 which is operational within 10 ms of stimulus onset. These authors differentiate this from the slow gain-control mechanism that takes effect over seconds, as reported by Albrecht, Farrar, and Hamilton (1984), Bonds (1991) and Ohzawa, Sclar, and Freeman (1985), amongst others. They argue that it is only the latter that may be useful for contrast discrimination, whereas the fast mechanism would only add greater uncertainty about changes in stimulus contrast occurring at the timescale of the psychophysical stimuli, but is instead useful for discrimination along other stimulus dimensions, such as orientation, spatial position, and spatial frequency. In sum, the sort of contrast adaptation that would desensitise neurons to high-contrast masks and therefore shift the c_{50} distribution to the right, as suggested by the third model's fits, is too slow to have been in operation in the psychophysical paradigm. In our experiments, as in the neurophysiology ones, high and low contrast would have been randomly interleaved so that the adaptational state would relate to the whole gamut of contrasts.

5. Conclusions

Our model of contrast discrimination has tried to incorporate three features of V1 physiology: the sigmoidal contrast–response functions of individual neurons each covering limited dynamic range, the multiplicative noise in the responses, and the differences in dynamic range location (c_{50}) of different neurons. Sampling the neuronal parameters for pools of neurons exactly as reported for V1 produced model dipper functions which were of approximately the correct form, and certainly much better than could be modelled with single neurons. However, a good fit to the psychophysical data required sacrifice of the true distribution of c_{50} in V1. The origins and significance of the systematic failures of the simple model (and of Boynton et al.'s (1999) comparison of fMRI and psychophysical data) require more investigation. It is intriguing that the distribution of c_{50} in V1 may be optimal for encoding the contrasts in natural scenes (Chirimuuta et al., 2003; Clatworthy et al., 2003) even though it does not immediately explain human contrast discrimination.

We have considered greater physiological detail than developed by previous analyses that have sought to understand the neural basis of the dipper function. An important finding has been that the well-known and elegant Naka–Rushton equation (Eq. (5)) is a poor basis for modelling contrast coding by V1, unless a small "hard threshold" is incorporated. Gorea and Sagi (2001) used a signal-detection-theory analysis of a novel masking experiment to infer that masking is due to a decelerating transducer function with additive noise. Kontsevich et al. (2002), on the other hand, argue that the effect is due to multiplicative noise in an accelerating channel. As it happens, our model integrates both explanations, showing how a neural system which saturates at high contrasts, *and* has multiplicative noise, will show masking effects. The classic models of Legge and Foley (1980) and Foley (1994) are empirical; they employ an abstract, "black-box" picture of the visual system and do not address the problem of how contrast is encoded neurally. The whole is modelled as if it were just a single neuron with a broad, non-linear transducer function and with additive response noise. In fact, this is essentially the same as the interpretation given to the fMRI BOLD signal by Boynton et al. (1999) and Heeger et al. (2000), essentially the "superneuron" of Shadlen et al. (1996). This summed neuronal signal is a good basis for predicting the dipper function, even without the additions of our model, i.e., distinguishing neuronal responses from each other and multiplicative noise. A Bayesian pooling rule retains information about differential neural firing rates, and should be the more efficient theoretically, but it is not known if such an operation takes place in brain. Clatworthy et al. (2003) showed that the Bayesian and the simple-summing pooling rules give qualitatively similar results, the performance of the Bayesian pooling rule just being a scaled up version of the other's. As regards the form of the noise: perhaps, some additive noise later in the decision process masks the multiplicative noise seen so obviously in V1 neurons.

The "non-linear transducer model" (Legge & Foley, 1980) and the "superneuron model" (Boynton et al., 1999) both provide explanations for the psychophysical dipper function. What, then, is the advantage of our Bayesian model of individual V1 neurons? By seeing inside the "black-box", we may gain insights into how the dipper function might change under different circumstances, for instance when mask and test take different orientations or spatial frequencies, or when the mask is more complex than a single grating (Chirimuuta & Tolhurst, 2004; Foley, 1994). The most important of our findings is that the *form* of the dipper (not just its location on the contrast axes) depends upon neuronal firing rates. For instance, if a masking stimulus were to reduce R_{max} of the detecting neurons (e.g., by non-specific suppression, Heeger, 1992a), this would not simply result in a rise in threshold; it would also result in a decrease in the depth of the dipper. Foley (1994) reported masking experiments with compound gratings that did produce results with shallower dips (a result not predicted by Legge & Foley (1980)), and he explained this with a psychophysical model involving non-specific suppression of the test channel by the mask channel. Our present modelling gives a clear intuition for this, and it raises the likelihood that the form of the dipper could change in different ways, for instance if some masking condition changed R_{max} of neurons with low c_{50} differently from neurons with high c_{50}.

Our simulations have also given a clue as to how many action potentials are generated in a population of neurons when high-contrast stimuli are presented; within the time interval during which a perceptual decision is made, the number of action potentials may be as few as 45, rising to a few hundreds (see legends to Figs. 7 and 9). These numbers (with their associated high variance) must pose limits on the precision with which populations of neurons encode visual information more generally (e.g., Tolhurst, Bulstrode, & Willmore, 2004).

Acknowledgments

This work was initiated while MC was supported by a MRC studentship in Cambridge, and was completed while MC was at Pisa, supported by the Consiglio Nazionale delle Ricerche, Italy. We are grateful to GT for acting as an observer. We thank the anonymous reviewers for their valuable comments. We also thank Dario Ringach for kindly providing us with neurophysiological data.

References

Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology, 347*, 713–739.

Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: Temporal dynamics of the contrast response function. *Journal of Neurophysiology, 88*, 888–913.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48*, 217–237.

Barlow, H. B., Kaushal, T. P., Hawken, M. J., & Parker, A. J. (1987). Human contrast discrimination and the threshold of cortical neurons. *Journal of the Optical Society of America A, 12*, 2367–2371.

Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *Journal of the Optical Society of America A, 19*, 1267–1273.

Bonds, A. B. (1991). Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience, 6*, 239–255.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39*, 257–269.

Boynton, G. M., & Foley, J. M. (1999). Temporal sensitivity of human luminance pattern mechanisms determined by masking with temporally modulated stimuli. *Vision Research, 39*, 1641–1656.

Bradley, A., & Ohzawa, I. (1986). A comparison of contrast detection and discrimination. *Vision Research, 26*, 991–997.

Campbell, F. W., & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology, 187*, 437–445.

Carandini, M., & Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience, 20*, 470–484.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*, 8621–8644.

Chirimuuta, M., Clatworthy, P. L., & Tolhurst, D. J. (2002). Simple-cell contrast responses and the transducer function. *Perception, 31*, 137A.

Chirimuuta, M., Clatworthy, P. L., & Tolhurst, D. J. (2003). Coding of the contrasts in natural images by visual cortex (V1) neurons: A Bayesian approach. *Journal of the Optical Society of America A, 20*, 1253–1260.

Chirimuuta, M., & Tolhurst, D. J. (2004). Contrast masking and facilitation in human psychophysical experiments using natural scene stimuli. *Journal of Physiology, 163*, 555P.

Chirimuuta, M., & Tolhurst, D. J. (2005). Accuracy of identification of grating contrast by human observers: Bayesian models of V1 contrast processing show correspondence between discrimination and identification performance. *Vision Research*, doi:10.1016/j.visres.2005.06.021.

Clatworthy, P. L., Chirimuuta, M., Lauritzen, J. S., & Tolhurst, D. J. (2003). Coding of the contrasts in natural images by populations of neurons in striate visual cortex (V1). *Vision Research, 43*, 1983–2001.

De Valois, R. L., Morgan, H., & Snodderly, D. M. (1974). Psychophysical studies of monkey vision III. Spatial luminance contrast sensitivity tests of macaque and human observers. *Vision Research, 14*, 75–81.

Dean, A. F. (1981). The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research, 44*, 437–440.

Foley, J. M. (1994). Human luminance pattern–vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, 11*, 1710–1719.

Gardner, J., Anzai, A., Ohzawa, I., & Freeman, R. (1999). Linear and nonlinear contributions to orientation tuning of simple cells in the cat's striate cortex. *Visual Neuroscience, 16*, 1115–1121.

Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. *Vision Research, 35*, 2723–2730.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience, 14*, 897–919.

Gorea, A., & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neuroscience, 4*, 1146–1150.

Gottschalk, A. (2002). Derivation of the visual contrast response function by maximizing information rate. *Neural Computation, 14*, 527–542.

Hawken, M. J., Parker, A. J., & Lund, J. S. (1988). Laminar organization and contrast sensitivity of direction selective cells in the striate cortex of the old-world monkey. *Journal of Neuroscience, 8*, 3541–3548.

Heeger, D. J. (1992a). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Heeger, D. J. (1992b). Half-squaring in responses of cat striate cortex. *Visual Neuroscience, 9*, 427–443.

Heeger, D. J., Huk, A. C., Geisler, W. S., & Albrecht, D. G. (2000). Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? *Nature Neuroscience, 3*, 631–633.

Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters—number, shapes and spatial covariation. *Vision Research, 32*, 47–59.

Itti, L., Koch, C., & Braun, J. (2000). Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America A, 17*, 1899–1917.

Kaplan, E., & Shapley, R. M. (1986). The primate retina contains 2 types of ganglion-cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America, 83*, 2755–2757.

Kiper, D. C., & Kiorpes, L. (1994). Sprathreshold contrast sensitivity in experimentally strabismic monkeys. *Vision Research, 34*, 1575–1583.

Kontsevich, L. L., Chen, C. C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42*, 1771–1784.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*, 149–162.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*, 1434–1448.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*, 1458–1470.

Lehky, S. R. (1985). Temporal properties of visual channels measured by masking. *Journal of the Optical Society of America A, 2*, 1260–1272.

Li, B. W., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology, 90*, 204–217.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology and perception. *Science, 240*, 740–749.

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain*. Cambridge, MA: MIT Press.

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4*, 930–943.

Nachmias, J., & Kocher, E. C. (1970). Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America, 60*, 382–389.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better that detection. *Vision Research, 14*, 1039–1042.

Naka, K. I., & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *Journal of Physiology, 185*, 536–555.

Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology, 42*, 833–849.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31*, 1337–1350.

Rao, R. P. N., Olshausen, B. A., & Lewicki, M. S. (2002). In *Models of the Brain*. Cambridge, MA: MIT Press.

Reich, D. S., Mechler, R., & Victor, J. D. (2001). Temporal coding of contrast in primary visual cortex: When, what and why. *Journal of Neurophysiology, 85*, 1039–1050.

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature Neuroscience, 6*, 414–420.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*, 281–284.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21*, 409–418.

Sclar, G., Maunsell, H. R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30*, 1–10.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioural responses to visual motion. *Journal of Neuroscience, 16*, 1486–1510.

Tolhurst, D. J. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *Journal of Physiology, 231*, 385–402.

Tolhurst, D. J., & Barfield, L. P. (1978). Interactions between spatial frequency channels. *Vision Research, 18*, 951–958.

Tolhurst, D. J., Bulstrode, H., & Willmore, B. (2004). Computer simulation of the effects of spike encoding on the representation of natural scene information in visual cortex. *Journal of Physiology, 555P*, C165.

Tolhurst, D. J., & Heeger, D. J. (1997). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience, 14*, 293–309.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research, 41*, 414–419.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23*, 775–785.

Vidyasagar, T. R., Kulikowski, J. J., Lipnicki, D. M., & Dreher, B. (2002). Convergence of parvocellular and magnocellular information channels in the primary visual cortex of the macaque. *European Journal of Neuroscience, 16*, 945–956.

Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research, 77*, 432–436.

Wiener, M. C., Oram, M. W., Liu, Z., & Richmond, B. J. (2001). Consistency of encoding in monkey visual cortex. *Journal of Neuroscience, 20*, 8210–8221.